

Análisis Exploratorio del Catálogo de Contenido de Netflix

Autor: Isaac Martinez Rios

Resumen

Este análisis exploratorio examina el catálogo de contenido de Netflix, con el objetivo de identificar patrones en el tipo de contenido, tendencias temporales, distribución por países y géneros predominantes. Los hallazgos revelan un catálogo dominado por películas, con una fuerte presencia de contenido estadounidense, y preferencia por géneros como dramas, comedias y películas internacionales. El análisis proporciona información valiosa para entender la estrategia de contenido de Netflix y las tendencias del mercado global de streaming.

Introducción

Netflix es una de las plataformas de streaming de vídeo y multimedia más populares a nivel mundial. A mediados de 2021, contaba con una biblioteca de más de 8.000 películas y programas de televisión, y una base de suscriptores que superaba los 200 millones en todo el mundo. Entender la composición de su catálogo proporciona valiosa información sobre tendencias de consumo, estrategias de adquisición de contenido y preferencias de audiencia en el mercado global de streaming.

Importancia del análisis

Este análisis exploratorio es relevante por múltiples razones:

- **Comprensión de estrategias comerciales:** Permite entender cómo Netflix selecciona y diversifica su contenido para atraer y retener suscriptores en diferentes mercados.
- **Identificación de tendencias globales:** Revela preferencias de consumo audiovisual a nivel mundial y cómo estas están evolucionando con el tiempo.
- **Oportunidades de mercado:** Ayuda a identificar nichos de contenido subrepresentados que podrían representar oportunidades para creadores y productores.
- **Perspectiva cultural:** Ofrece una visión de cómo la globalización del entretenimiento está influenciando los hábitos de consumo de contenido.

Descripción de los Datos

Este análisis utiliza un conjunto de datos obtenido de [Kaggle](#), que contiene información detallada sobre todas las películas y programas de televisión disponibles en Netflix. El dataset incluye 8.807 entradas con 12 columnas de información.

Estructura del Dataset:

"show_id": #Identificador único de la serie o película.
"type": #Indica si el contenido es una película o serie de TV.
"title": #Título del programa o película.
"director": #Nombre del director del programa o película.

"cast": #Lista de actores presentes en el programa o película.
"country": #País donde se produjo el programa o película.
"date_added": #Fecha en que el programa o película fue añadido a Netflix.
"release_year": #Año en el que se lanzó el programa o película.
"rating": #Clasificación del contenido del programa o película.
"duration": #Duración del programa o película.
"listed_in": #Categoría a la que pertenece el programa o película.
"description": #Descripción del programa o película.

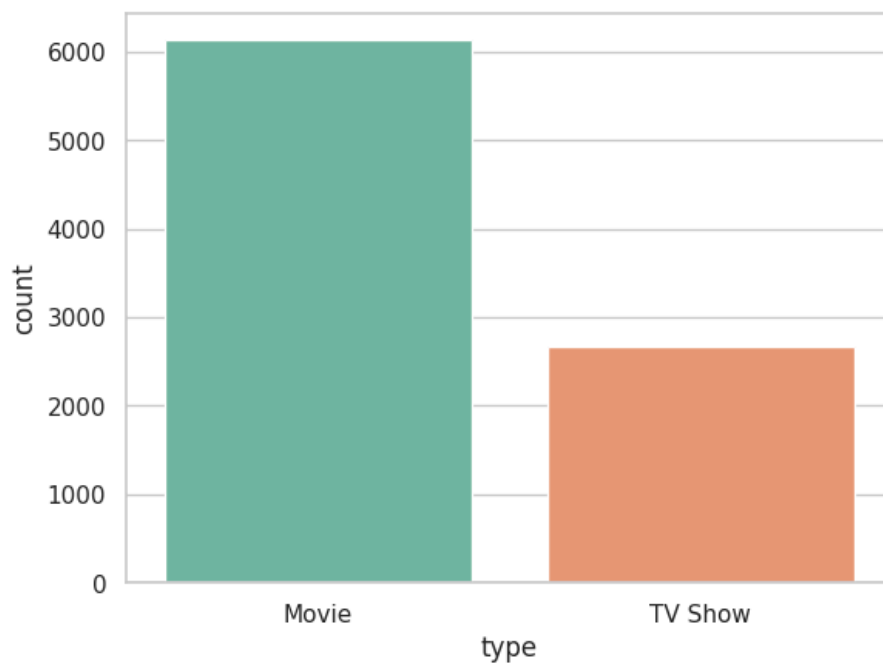
Calidad de los Datos:

```
RangeIndex: 8807 entries, 0 to 8806  
Data columns (total 12 columns):  
#   Column      Non-Null Count  Dtype  
---  ---  
0   show_id     8807 non-null   object  
1   type        8807 non-null   object  
2   title       8807 non-null   object  
3   director    6173 non-null   object  
4   cast        7982 non-null   object  
5   country     7976 non-null   object  
6   date_added  8797 non-null   object  
7   release_year 8807 non-null   int64  
8   rating      8803 non-null   object  
9   duration    8804 non-null   object  
10  listed_in   8807 non-null   object  
11  description 8807 non-null   object
```

Se observa que algunas columnas contienen valores nulos, especialmente `director` (2,634 valores faltantes), `cast` (825 valores faltantes) y `country` (831 valores faltantes), lo que debe considerarse durante el análisis.

Análisis Exploratorio de Datos (EDA)

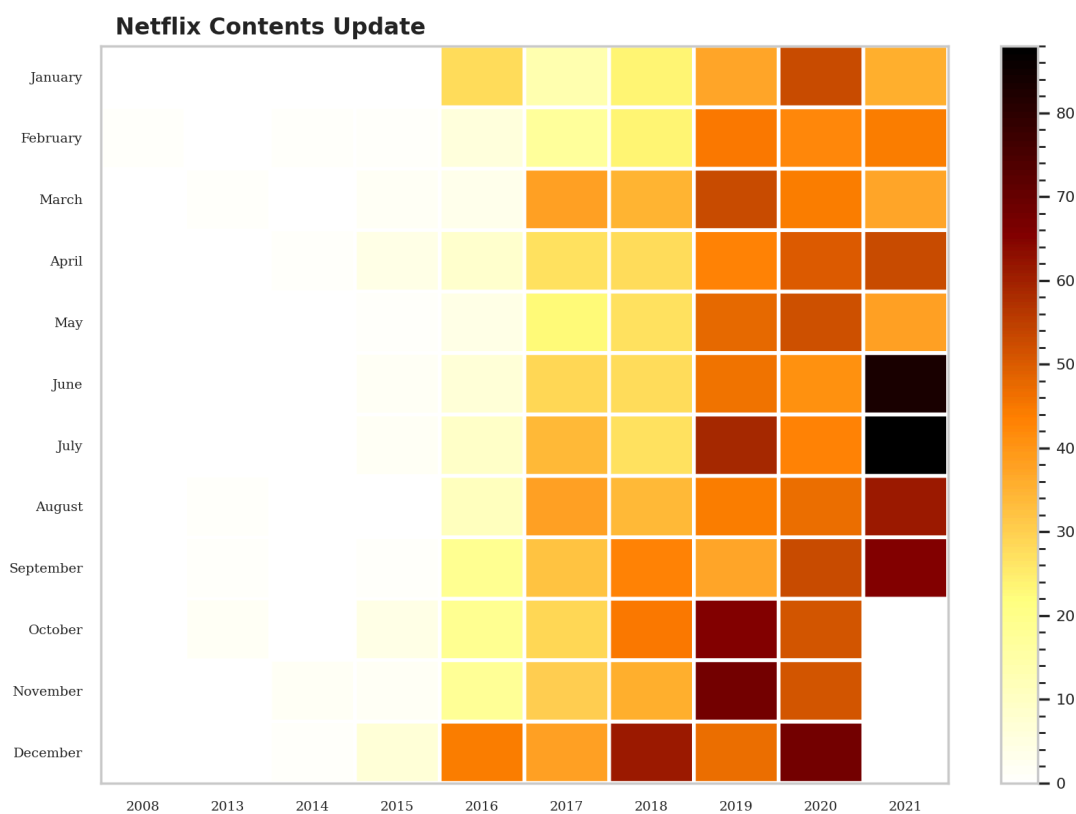
Análisis de películas frente a programas de televisión



Observaciones

El gráfico muestra la distribución de contenido entre películas y series de TV en el catálogo de Netflix. Esto nos permite entender el enfoque de adquisición de contenido de la plataforma.

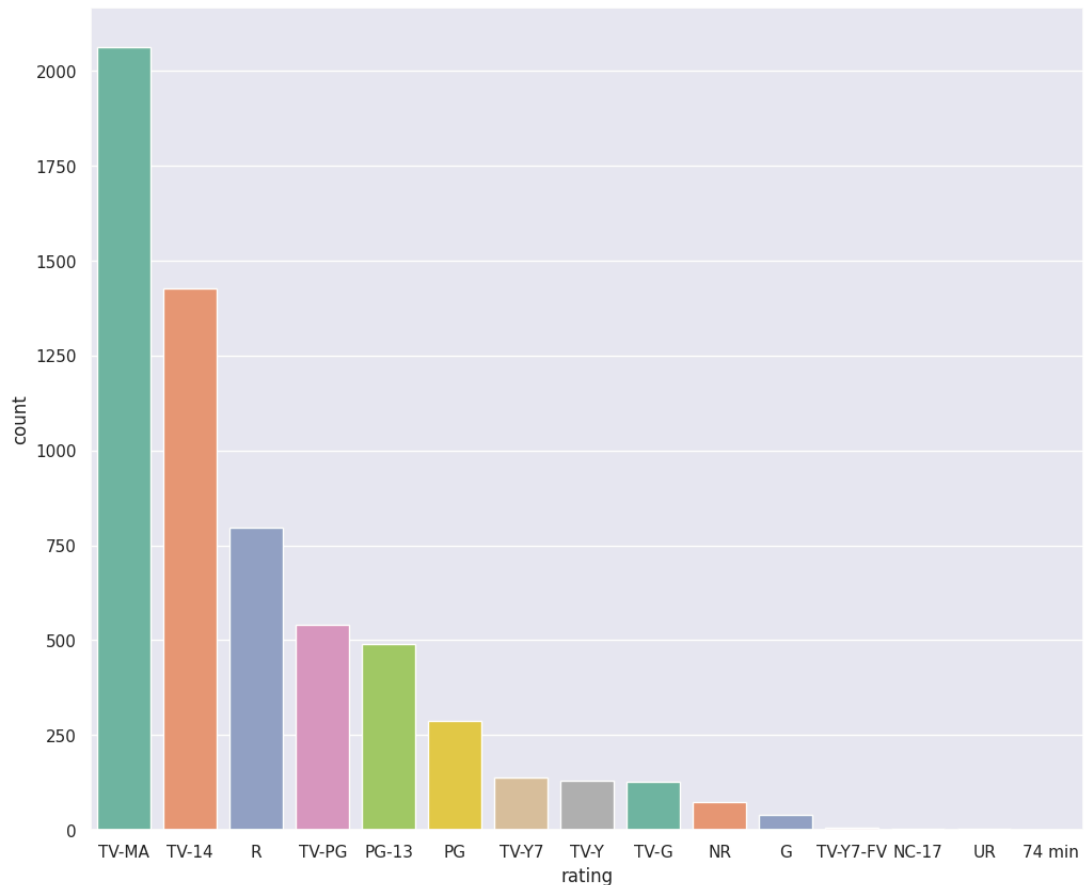
Actualización de contenido de Netflix



Observaciones

Este análisis temporal muestra cómo Netflix ha ido añadiendo contenido a su plataforma a lo largo del tiempo, permitiéndonos identificar períodos de mayor adquisición de contenido.

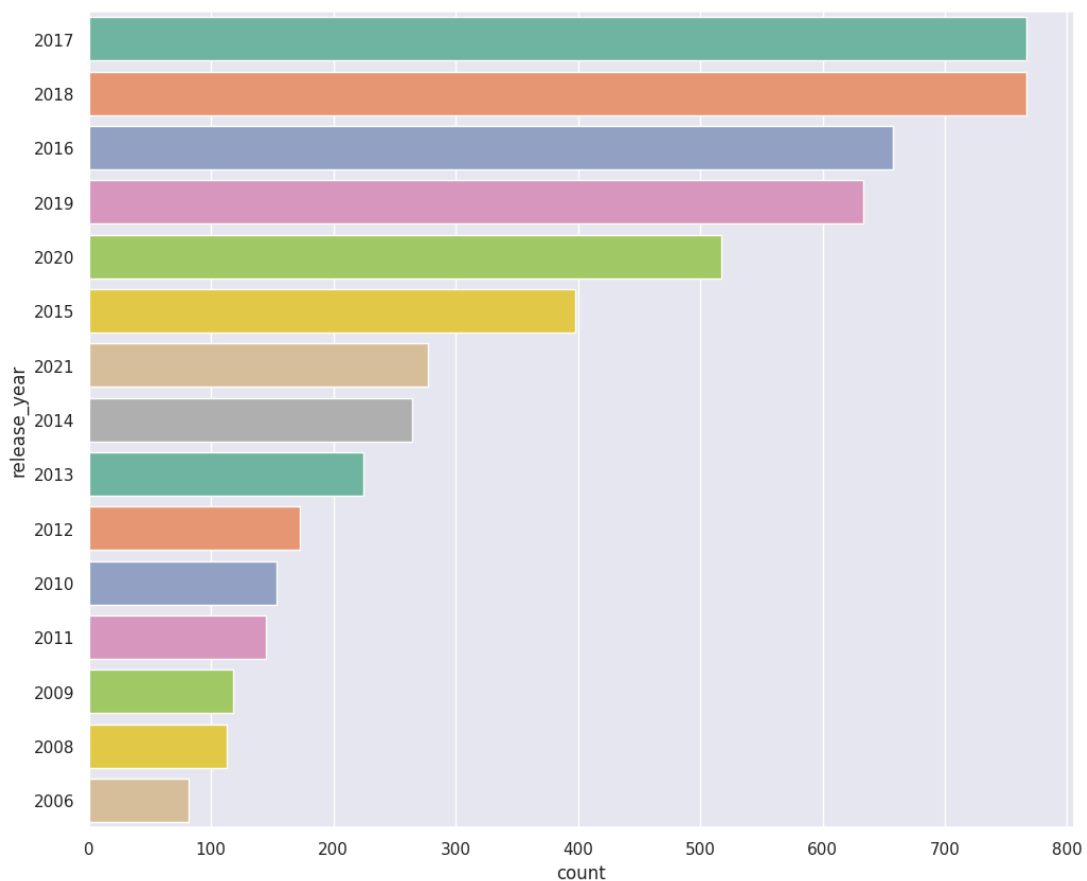
Análisis de calificaciones de películas



Observaciones

Esta distribución refleja una estrategia de contenido claramente orientada hacia audiencias adultas y adolescentes mayores, con un enfoque secundario en contenido familiar. El predominio de clasificaciones TV-MA, TV-14 y R sugiere que Netflix ha apostado principalmente por atraer a suscriptores adultos con contenido más maduro y complejo.

Análisis anual de estrenos



Observaciones

- El período 2016-2019 muestra un crecimiento sustancial en la cantidad de contenido añadido, con 2018 como el año con mayor número de estrenos (aproximadamente 780 títulos).
- Se observa una clara evolución ascendente desde 2006 hasta 2018, seguida de una ligera disminución en 2019 y una caída más pronunciada en 2020 (posiblemente relacionada con la pandemia de COVID-19).

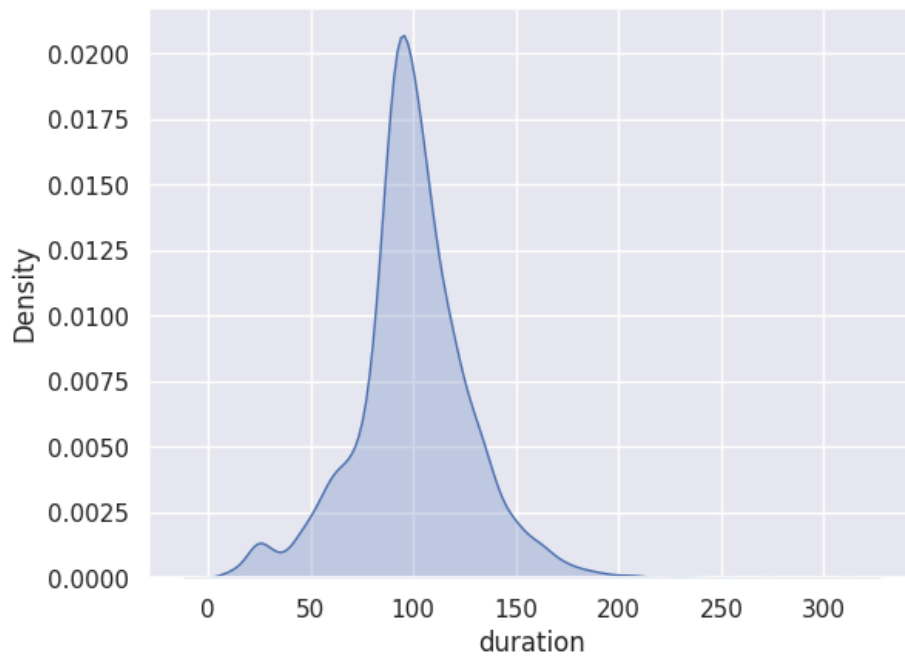
Períodos identificables:

- **2006-2013:** Fase de crecimiento inicial y moderado, con menos de 300 títulos por año.
- **2014-2015:** Fase de transición con aumento notable en adquisiciones.
- **2016-2019:** Período de expansión agresiva, con más de 600 títulos añadidos anualmente.
- **2020-2021:** Fase de contracción o estabilización.

Implicaciones:

- El pico en 2018 coincide con la estrategia de Netflix de expandir significativamente su biblioteca de contenido original y adquirido.

Análisis de duración de películas

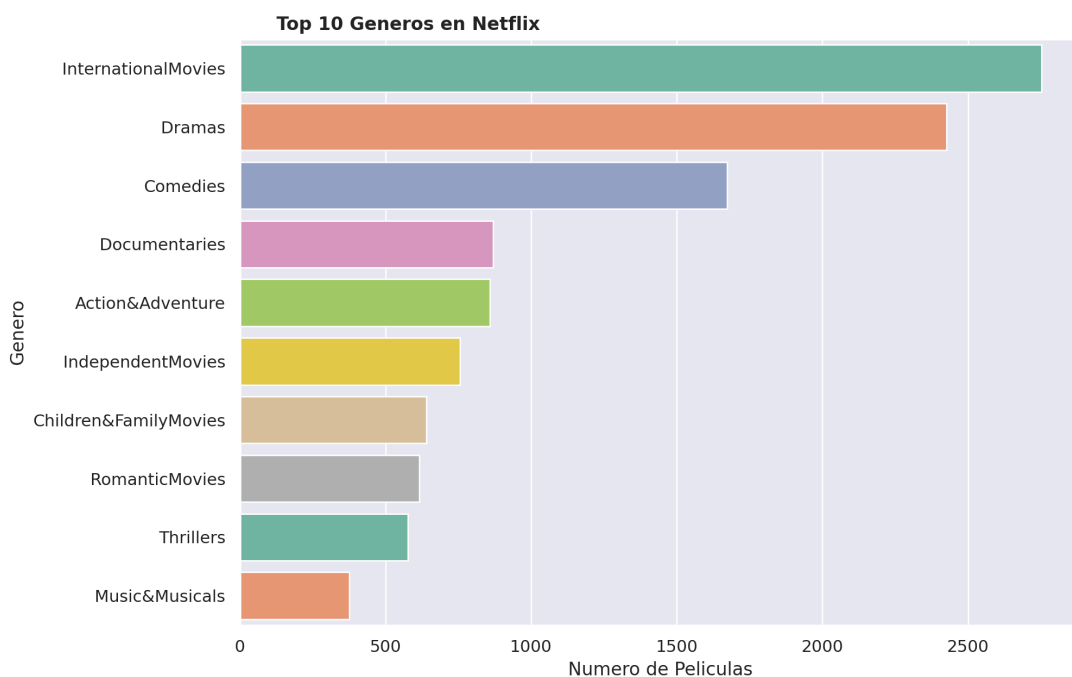


Observaciones

- **Punto máximo:** La mayor concentración de películas se sitúa alrededor de los 90-110 minutos de duración.
- **Rango principal:** La mayoría de las películas tienen una duración entre 75 y 120 minutos.

Netflix prioriza contenido de duración media, lo cual es coherente con los hábitos de consumo de los espectadores que prefieren contenido que pueda verse en una sola sesión.

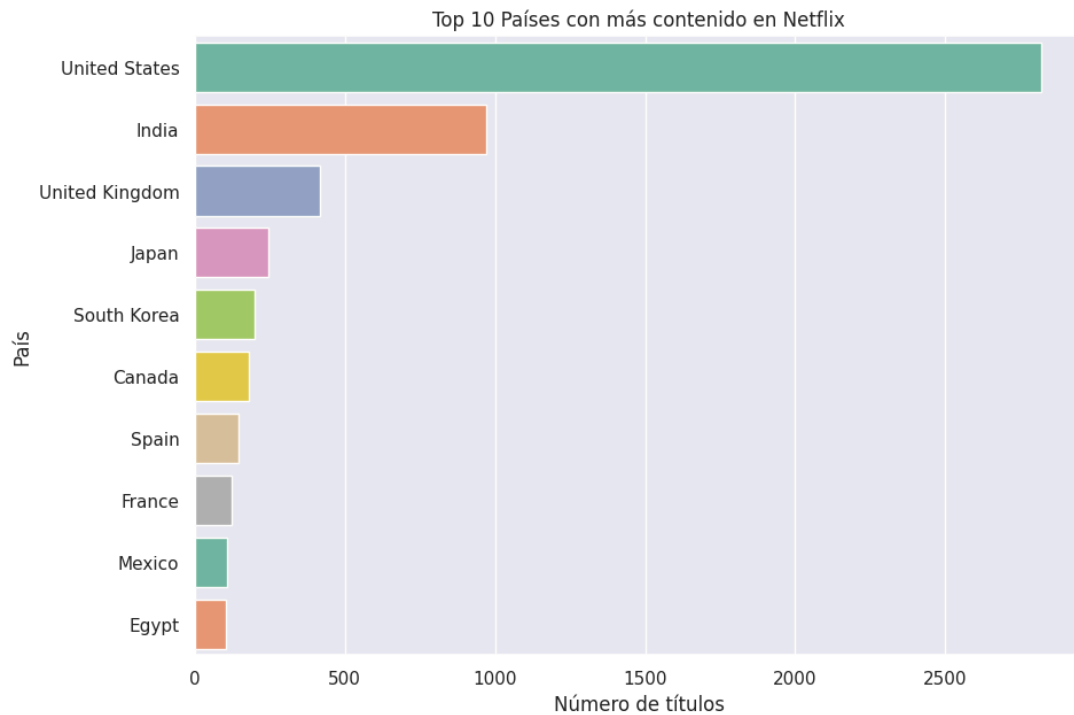
Géneros más populares en Netflix



Observaciones

Podemos ver que los géneros predominantes (internacionales, dramas y comedias), y muestra la estrategia de contenido de Netflix orientada a satisfacer preferencias globales diversas, con especial énfasis en contenido internacional.

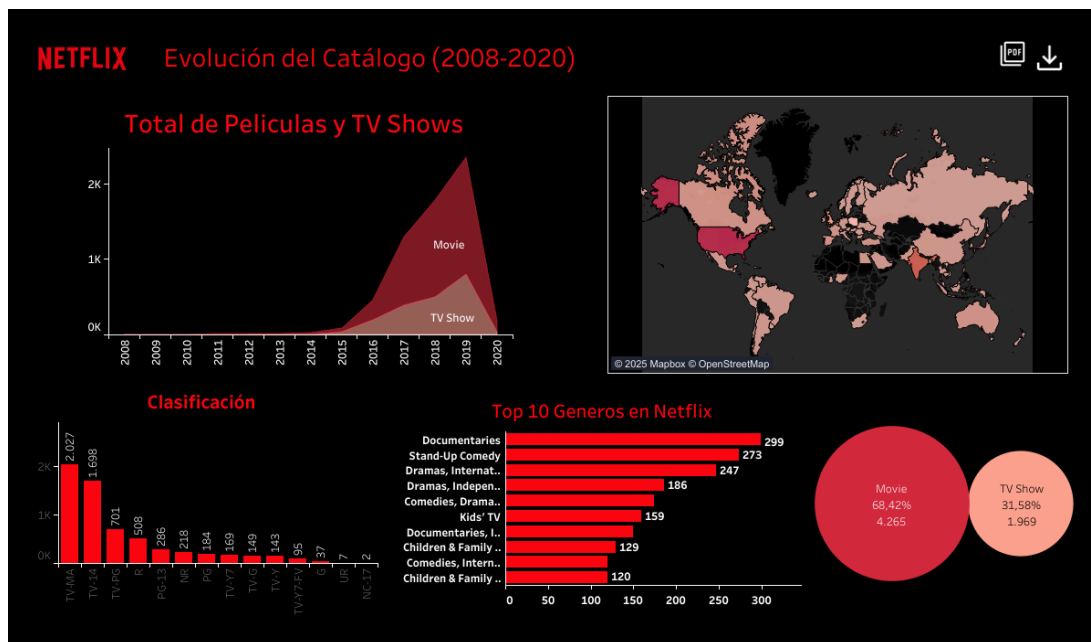
Países que producen la mayor cantidad de contenido



Observaciones

El gráfico revela un notable desequilibrio en la distribución geográfica del contenido, con una fuerte concentración en pocos países, principalmente Estados Unidos e India. Esta distribución refleja tanto las estrategias de adquisición de contenido de Netflix como las dinámicas de la industria global de producción audiovisual.

Visualización de Datos en Tableau



https://public.tableau.com/views/Netflix_Dashboard_17405133855050/NetflixDashboard?:language=es-ES&:sid=&:redirect=auth&:display_count=n&:origin=viz_share_link

Hallazgos y Discusión

A partir del análisis exploratorio, podemos destacar los siguientes hallazgos clave:

- **Dominancia de películas:** El catálogo de Netflix presenta una mayor cantidad de películas que series, lo que refleja su estrategia de adquisición de contenido.
- **Duración óptima:** Las películas más comunes tienen una duración entre 75 y 120 minutos, alineándose con las preferencias de visualización de los usuarios.
- **Géneros predominantes:** El contenido internacional, dramas y comedias dominan el catálogo, indicando los gustos globales del público.
- **Hegemonía estadounidense:** Estados Unidos produce la mayor parte del contenido, seguido por India y Reino Unido, reflejando el poder de estas industrias cinematográficas.
- **Diversificación regional:** Aunque existe representación de múltiples regiones (Asia, Europa, Latinoamérica, África), hay un claro desequilibrio en favor de pocos países.

Conclusiones

- **Netflix ha construido un catálogo diverso pero desequilibrado:** Aunque incluye contenido de diferentes regiones, existe una clara predominancia de producciones estadounidenses.
- **La plataforma prioriza películas sobre series:** Esto podría responder a una estrategia de costo-beneficio o a preferencias de la audiencia.
- **El contenido reciente predomina:** Indicando una constante actualización del catálogo para mantener su relevancia.

Apéndices

https://github.com/isaacmartinez23/portafolio-Data-Analyst/blob/main/EDA_Netflix.ipynb

<https://www.kaggle.com/datasets/shivamb/netflix-shows>